Federated Adaptive Pseudo-labeling Selection for Semi-Supervised Air Writing Recognition Systems

Wenyi Zhang, Xiangjie Kong, *Senior Member, IEEE,* Guojiang Shen, Zhenpeng Wu, and Youyang Qu, *Member, IEEE*

***Abstract*—The rapid advancement of virtual and augmented reality technologies has catalyzed a new paradigm in consumer electronics: air writing, a form of touchless human-computer interaction with vast potential. Implementing air writing recognition systems, however, faces challenges such as label scarcity and privacy concerns. Addressing these, we propose the Federated Adaptive Pseudo-labeling Selection (FedAPS) framework, a federated semi-supervised learning approach, enhancing decision-making in consumer electronics using multi-modal data. FedAPS innovatively utilizes limited labeled data alongside extensive unlabeled data, maintaining user privacy. We employ a multi-modal data augmentation process for air writing recognition by designing an adaptive pseudo-labeling strategy. This enables clients to select the most appropriate model for pseudo-labeling based on historical local and global models and dynamic word score recommendations. We introduce a historical local-global consistency regularization to maximize knowledge extraction from unselected models when similar predictions occur. Our comprehensive evaluation on a real-world multi-modal air writing dataset shows FedAPS's effectiveness, outperforming advanced federated semi-supervised baselines and achieving performance comparable to fully labeled federated supervised learning. This highlights its potential in enhancing data-driven decision-making for next-generation multi-modal input consumer electronics.**

***Index Terms*—Federated adaptive learning, air-writing electronics, semi-supervised learning, pseudo-labeling, multi-modal based text recognition.**

I. Introduction

TEXT entry serves as a fundamental element in human-computer interaction applications for consumer electronics [1]. As the field of human-computer interaction continues to evolve, there is a growing interest in multi-modal human-computer interaction, which integrates various forms of interaction such as keyboard, sensors [2], speech [3] and gestures [4], as shown in Fig. 1. This technology has attracted the growing attention of both academia and industry due to its potential to introduce new possibilities for text entry and consumer electronics.

Physical keyboards dominated the field during the early stages of human-computer interaction research, given their intuitive nature [1]. However, with the advancement of mobile computing, soft keyboards were developed to enhance mobility [5]. More recently, virtual or invisible keyboards have been introduced to maximize the practicality of device screens [6]. In addition, speech input has gained significant attention and usage. By combining technologies such as speech recognition, gesture recognition, and eye tracking, users can interact with devices in a more natural way, enhancing convenience and personalization. In the area of text entry, multi-modal human-computer interaction technology offers users more convenient and efficient entry methods, effectively overcoming the limitations of traditional keyboard entry. At the same time, this technology enables consumer electronics products to deliver more intelligent and more user-friendly interactive experiences, enabling products with diverse functionalities and application scenarios [7].

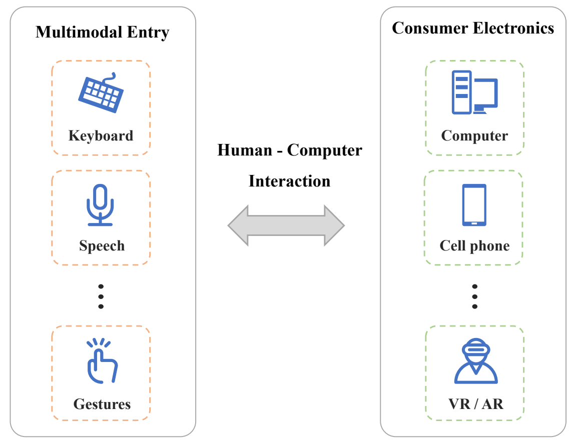


Fig. 1. Multi-modal human-computer interaction in consumer electronics.

Recent advancements in Virtual Reality (VR) and Augmented Reality (AR) technologies have propelled the development of new paradigms in contactless human-computer interaction [1]. The prospect of air-writing, a novel mode of interaction within this technological landscape, holds immense promise for the consumer electronics industry, especially the ones with multi-modal data. This technology allows users to interact with digital interfaces through gestures and handwriting in the air, presenting a futuristic and intuitive mode of interaction. This has sparked considerable interest among researchers and research teams, driving a surge in studies to explore the potential applications of air writing [1], [8]–[10].

However, despite the considerable strides in enhancing the

This work was supported in part by the National Natural Science Foundation of China under Grant 62072409 and Grant 62073295, and in part by the Zhejiang Provincial Natural Science Foundation under Grant LR21F020003.

Wenyi Zhang, Xiangjie Kong, Guojiang Shen and Zhenpeng Wu are with the College of Computer Science and Technology, Zhejiang University of Technology, Hangzhou 310023, China. E-mail: wyzhang98@outlook.com, xjkong@ieee.org, gjshen1975@zjut.edu.cn, wzpzzz6@outlook.com.

Youyang Qu is with data61, Australia Commonwealth Scientific and Industrial Research Organization (CSIRO), Australia. E-mail: youyang.qu@data61.csiro.au.

Corresponding author: Xiangjie Kong

accuracy of multi-modal air writing recognition models, the predominant focus has been on leveraging vast amounts of labeled data [11]–[13]. The practical challenge lies in the scarcity of labeled air-writing data, posing a significant impediment to the widespread deployment and practical application of air-writing in consumer electronics. Issues such as label scarcity, user privacy concerns, and the need for efficient models that can operate with limited labeled data become apparent in this context [8], [14], [15]. In addition, the imperative need to protect user privacy in the era of pervasive multi-modal data collection further complicates the development and deployment of such systems [16], [17].

In response to these challenges, Federated Semi-Supervised Learning (FSSL) has been introduced as a promising avenue to overcome the constraints imposed by the limited availability of labeled multi-modal data while respecting user privacy. FSSL enables multiple client devices to collaboratively leverage labeled and unlabeled multi-modal data from all participating clients without centralized data exchange [18]. Despite the potential of semi-supervised learning in a federated setting, the direct combination of semi-supervised learning methods with federated learning proves sub-optimal [19]. Existing FSSL approaches exhibit shortcomings that hinder their efficacy in scenarios like air writing recognition.

Our proposed framework, FedAPS, operates within the domain of FSSL, a paradigm that leverages labeled and unlabeled multi-modal data to enhance model generalization. The objective is to empower air writing recognition models with improved adaptability and accuracy, even when labeled multi-modal data is limited. This innovative framework is specifically tailored for air writing recognition scenarios, addressing the limitations of existing FSSL methods. FedAPS incorporates an adaptive pseudo-labeling strategy to achieve multi-modal data augmentation. This enables client devices to autonomously choose the most appropriate model to generate pseudo-labels for unlabeled multi-modal data. The adaptive strategy considers historical local models, global models, and recommendations from dynamic word score algorithms, thereby significantly enhancing the efficient augmentation from unlabeled data to well-labeled multi-modal data. Furthermore, we introduce a historical local-global consistency regularization term, which enhances the synergy between historical local and global models to maximize knowledge exploitation. The main contributions of this paper are as follows.

- We introduce an adaptive pseudo-label selection federated semi-supervised learning framework tailored for air writing recognition scenarios considering multi-model data input.
- We propose an adaptive pseudo-label selection strategy for multi-modal data augmentation, which leverages the selection of historical local models, global models, and recommendation algorithms to generate pseudo-labels for unlabeled data, thereby making more efficient use of unlabeled data.
- Experimental evaluations conducted on a real air writing multi-modal dataset demonstrate superior performance compared to various federated semi-supervised baselines and comparable performance to federated supervised learning with complete labels.

The rest of the paper is structured as follows. Section 2 presents a review of the relevant work. Section 3 introduces the proposed federated learning framework with selective pseudo-labeling for semi-supervised air writing recognition. Section 4 provides a thorough evaluation of the FedAPS framework on a publicly available air writing dataset. Lastly, Section 5 concludes the paper and outlines directions for future research.

II. Related Work

In this section, we review the related work in air writing recognition, semi-supervised learning, and federated semi-supervised learning.

A. Air Writing Recognition

Air writing, as one of the futuristic human-computer interaction methods, has garnered significant attention. Users can naturally write in the air with their fingers, and the Air Writing Recognition (AWR) system interprets the trajectory of the user's handwriting, converting it into the intended textual content [1]. In recent years, various research groups have extensively studied this field [1], [8]–[10], [13]. Most existing AWR methods predominantly rely on fully supervised deep learning models. However, during the learning process, these models often require a substantial amount of high-quality labeled data, which proves to be challenging. As a result, the efficient deployment of AWR systems in real-world scenarios remains constrained by various limitations and challenges. Key issues encompass the scarcity of labeled data, the necessity for model personalization across different users, and considerations related to user privacy [8]. This study specifically addresses the challenge of scarce labeled multi-modal data in training AWR models for multiple users. Resolving this issue holds significant importance for enhancing the performance of AWR systems and promoting their widespread application in consumer electronics.

B. Semi-supervised Learning

Achieving high performance in models for air writing recognition tasks requires abundant labeled multi-modal data. However, obtaining such labeled multi-modal data is often challenging and expensive [8]. Various approaches have been explored to address the scarcity of labeled data in machine learning, including pseudo-labeling [20]–[22] and consistency regularization [23], [24]. Pseudo-labeling, a popular method, simplifies the acquisition of labeled data by applying a threshold to the predicted maximum probability, thus assigning pseudo-labels to data samples [20]. This technique has been extended to dynamic thresholds based on the correlation between samples and labeled data [22] and category-based thresholds [21]. Another commonly used approach is consistency regularization, which exploits unlabeled data by encouraging models to generate consistent predictions under input perturbations [23], [24]. Recent research has combined these techniques [25], [26], but these efforts have primarily focused on centralized training, overlooking challenges arising from the limited generalizability of labeled data to unlabeled data and the complexity introduced by heterogeneous data across clients [27]. Moreover, integrating semi-supervised learning methods directly into the federated learning framework has proven suboptimal [19]. Our proposed FedAPS method demonstrates robustness to limited labeled multi-modal data and data heterogeneity in this context.

C. Federated Semi-supervised Learning

Recently, federated semi-supervised learning has gained significant traction in the research community [27]–[29], especially in addressing more practical federated learning scenarios, which can be mainly classified into three typical scenarios. The first two involve situations where labeled data is accessible only on a central server or some clients [28], [30], leaving the remaining clients with unlabeled data [19], [31]–[33]. Our focus lies on the third scenario, where each client has limited labeled data and a substantial amount of unlabeled data. Our work is closely related to FedMatch [32], FedTriNet [33], and FedLabel [19]. In FedMatch, the server identifies clients with similar data distributions and sends predictions from those clients to achieve inter-client consistency regularization. FedTriNet employs three networks and dynamic quality control to generate high-quality pseudo-labels for unlabeled data, incorporating them into the training set. However, these methods come with additional communication and computational costs, and their performance depends heavily on the number of assisting clients or how the models are concatenated [19]. FedLabel generates pseudo-labels through binary selection between local and global models, potentially losing valuable pseudo-label knowledge when neither provides accurate predictions. These approaches show limited effectiveness when the percentage of labeled data is low, and data heterogeneity is high.

However, existing methods barely consider the multi-modal air writing problem, along with limited multi-modal data, privacy, performance, etc. The proposal of FedAPS aims to address the above issues and to advance the knowledge of multi-modal air writing systems.

III. Methodology and System Modelling

In this section, we first introduce the problem formulation of FedAPS over the unlabeled data of clients, followed by a detailed explanation of the overarching algorithm framework of FedAPS, explicitly tailored to be implemented for AWR tasks in real-world FL scenarios. An overview of the FedAPS workflow is depicted in Fig. 2.

A. Problem Formulation

In practical FSSL scenarios, the client typically faces a situation where there are a few labeled data alongside an abundance of unlabeled data. The ability to generalize from labeled data to extensive unlabeled data may be limited due to factors such as the mismatch in class distributions between $\mathcal{D}_{L,k}$ and $\mathcal{D}_{U,k}$, as well as the inherent limitation imposed by the scarcity of labeled samples. In addition, Further compounding the challenge of FSSL is the presence of data heterogeneity, characterized by the discrepancy in the distributions of local data $\mathcal{D}_k$ amongst different clients. This work shows how our proposed FedAPS approach empowers clients to selectively utilize the global model, historical local model, and dynamic word score recommendation to exploit unlabeled data effectively. This is achieved despite the challenges posed by data heterogeneity and the limited generalizability of labeled data, as extensively discussed in the following section.

B. The Proposed FedAPS

In the prevalent scenarios of vanilla federated learning, clients have access to two primary sources of information for learning: the global model and the local model. The effectiveness of labeling the unlabeled data relies on which model, whether local, global, or even both, has a superior comprehension of the data based on their learned knowledge. If the client possesses only a fraction of labeled data, resulting in a class distribution mismatch or restricted capacity to generalize to unlabeled data. In such cases, it may be difficult for the client to provide accurate pseudo-labeling for the unlabeled data. However, the global model is more likely to produce appropriate labels for unlabeled data as it has been trained across different clients and has seen more data from various clients. If the local labeled data from the client is sufficient to make reliable generalizations to the local unlabeled data, then the local model is more likely to produce accurate pseudo-labels.

*1) **Acquiring Models and Initialization**:* The superscripts (t, r) represent the communication round t and the local iteration r, respectively. For each t, the server randomly selects some clients $\mathcal{C}^t$ and distributes the globally aggregated global model $w^{(t,0)}$ to these selected clients. The clients in $\mathcal{C}^t$ initialize the global model $w_{\mathcal{G},k}^{(t,0)} = w^{(t,0)}$, the historical local model $w_{\mathcal{H},k}^{(t,0)} = w_{\mathcal{L},k}^{(t-1,\tau)}$ saved from the previous round, and the local model $w_{\mathcal{L},k}^{(t,0)} = w^{(t,0)}$ to be updated iteratively for the current round.

*2) **Binary Selection based Pseudo-labeling**:* For each client $k \in \mathcal{C}^t$, the probability vectors from each global and historical local model are obtained from the unlabeled data $u \in \mathcal{D}_{U,k}$, denoted as $p(w_{\mathcal{G},k}^{(t,0)}, u)$ and $p(w_{\mathcal{H},k}^{(t,0)}, u)$, respectively, where $p(\cdot,\cdot) : \mathbb{R}^q \times \mathbb{R}^d \longrightarrow \mathbb{R}^{N\times M}$. Subsequently, we use the function $s(\cdot) : \mathbb{R}^{N\times M} \longrightarrow \mathbb{R}$ to compute the confidence score of the probability vector, i.e., the multiplication of the highest logical probabilities for each letter position based on the greedy algorithm. The word associated with the highest confidence score is chosen as a potential candidate for pseudo-labeling. Formally, as follows:

Calculate the Confidence Score:

$$s(p) = \prod_{i}^{N} \max_{j} p_{i,j}, \tag{1}$$

Selection of Probability Vector:

$$\begin{aligned} p^*(u) &= \arg\max_{p\in\mathcal{P}(u)} s(p), \\ \mathcal{P}(u) &:= \{p(w_{\mathcal{G}}, u), p(w_{\mathcal{H}}, u)\}, \end{aligned} \tag{2}$$

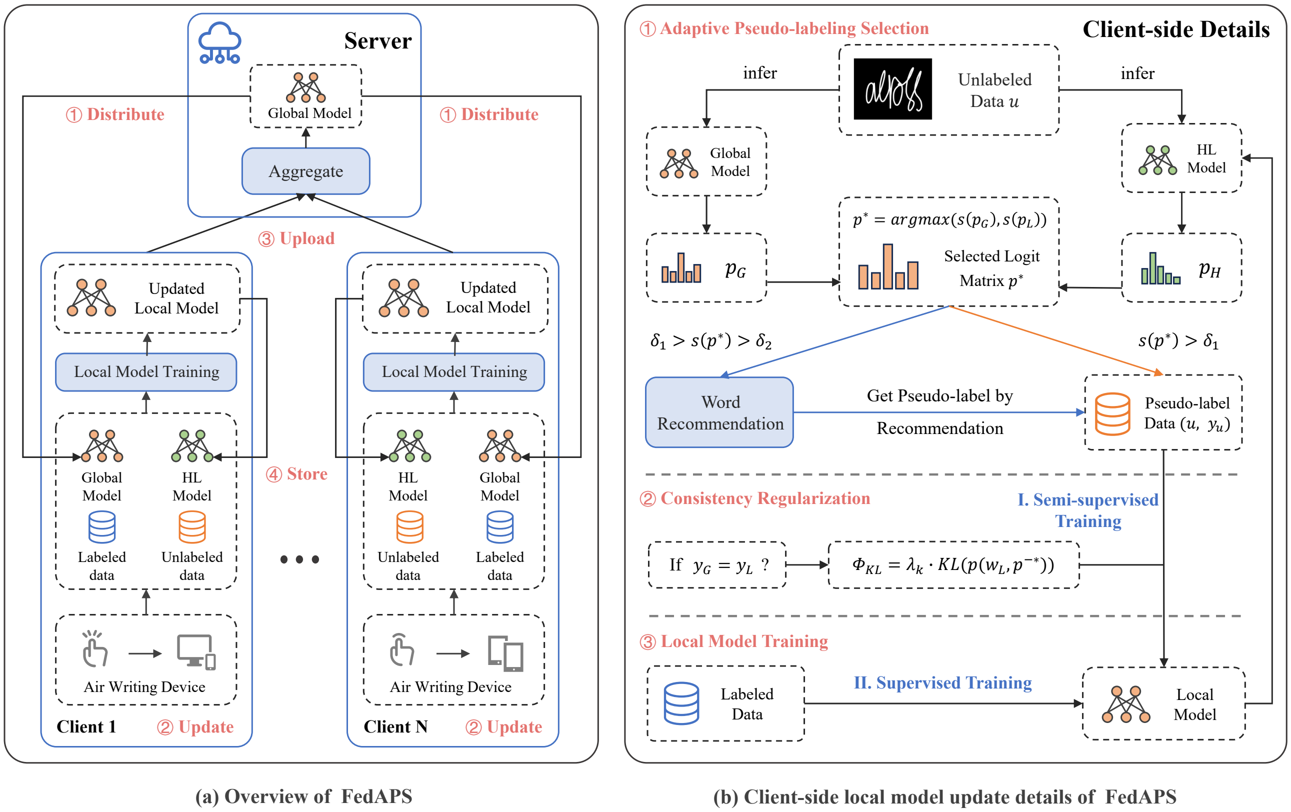


Fig. 2. The overview of the FedAPS framework and its client-side local model update details.

Pseudo-label via Threshold:

$$\hat{y}_u = \arg\max_j p^*_{i,j} \quad \text{s.t.} \quad s(p^*) > \delta_1. \tag{3}$$

If the confidence score of the selected probability matrix passes the threshold δ_1, we can obtain the pseudo-label $\hat{y}_u$ for unlabeled data $u \in \mathcal{D}_{U,k}$ via Eq. 3.

*3) **Recommendation based Pseudo-labeling**:* However, if the confidence score of the chosen model falls below the predetermined threshold δ_1, we will discard that unlabeled data. This inevitably results in potentially losing valuable information from this part of unlabeled data. To maximize the utilization of unlabeled data and cover more situations, we propose a pseudo-labeling approach based on dynamic word score recommendation inspired by AirText [10] for the air writing scenario. The approach consists of four primary steps, i.e., inferring soft labels, recommending candidate words, calculating dynamic word scores, and filtering pseudo-labels.

The first step is inferring soft labels. For each client $k \in \mathcal{C}^t$, the soft labels denoted $\hat{y}_\mathcal{G}$ and $\hat{y}_\mathcal{H}$, respectively, come from each global model and historical local model, obtained from their corresponding probability vectors p by the following formula:

$$\hat{y} = \arg\max_j p_{i,j}. \tag{4}$$

The second step is recommending candidate words. The recommendation operation is performed on its soft label solely when the confidence score of the model's probability matrix exceeds the recommendation threshold, i.e., $s(p) > \delta_2$. We then use $WR(\cdot)$ to provide word recommendations for soft labels based on the Damerau-Levenshtein distance. Additionally, a trie tree data structure is designed to store the word frequencies to speed up searching for words and their word frequencies based on the Damerau-Levenshtein distance from the linguistic database. Incrementally increasing the Damerau-Levenshtein distance until five candidate words from the language database most closely resemble the soft-label words are retrieved, formally as follows:

$$\mathcal{W} := WR(\hat{y}) \quad \text{s.t.} \quad s(p) > \delta_2, \tag{5}$$

where $\mathcal{W}$ is the set of candidate words and their word frequencies. By combining the $\mathcal{W}_G$ and $\mathcal{W}_H$ from the global model with the historical local model, we have derived the set of ultimate candidate words $\mathcal{W}_R$ as displayed below:

$$\mathcal{W}_R := \mathcal{W}_\mathcal{G} \cup \mathcal{W}_\mathcal{H}. \tag{6}$$

The third step is calculating dynamic word scores. Consider $w' = c_1 c_2 \cdots c_N, w' \in \mathcal{W}_R$ as the candidate word, where N is the length of the word. The dynamic word score $h(w')$ of each candidate word may then be determined by applying the following formula:

$$h(w') = \mathcal{F}(w') \cdot \prod_i^N p_{i,j}(c_i), \tag{7}$$

where $p_{i,j}(c_i)$ represents the logical probability of character c_i at the i-th position of word w', which is the dynamic component, $\mathcal{F}(w')$ denotes the word frequency of word w' in the linguistic database, serving as the static component. With this combination of dynamic and static calculations, higher dynamic word scores mean that they are not only dynamically closer to the model's predictions but are also statically higher frequency words.

The fourth step is filtering pseudo-labels. Pseudo-labeling by selecting the candidate word with the maximum dynamic word score among the candidates:

$$\hat{y}_u = \arg\max_{w' \in \mathcal{W}_R} h(w'). \tag{8}$$

Combining the binary selection based pseudo-labeling from Eq. 3 and the recommendation based pseudo-labeling from Eq. 8, the final pseudo-labeling $\hat{y}_u$ definition can be expressed as:

$$\hat{y}_u = \begin{cases} \arg\max_{j} p^*_{i,j}\,, & s(p^*) > \delta_1; \\ \arg\max_{w' \in \mathcal{W}_R} h(w')\,, & s(p^*) > \delta_2 \,||\, s(p^{-*}) > \delta_2. \end{cases} \tag{9}$$

Then, we can utilize the local model separately setting for this round, denoted as $w^{(t,0)}_{\mathcal{L},k}$, to train it on unlabeled data with pseudo-labels. In this case, the cross-entropy loss on unlabeled data for FedAPS can be expressed as:

$$\Phi_{CE}(u) = CrossEntropy(w^{(t,0)}_{\mathcal{L},k}, u, \hat{y}_u). \tag{10}$$

*4) **Local-Global Consistency Regularization**:* When the selected model's confidence score $s(p^*(u))$ exceeds the threshold δ_1, FedAPS calculates the cross-entropy loss in Eq. 10 using the pseudo-label selected via a binary selection process between the historical local and global models. This approach intrinsically discards the logical probability matrix of the non-selected model denoted as $p^{-*}(u)$.

However, there is a scenario where the soft labels predicted by the unselected model are identical to those of the selected model despite the lower confidence score for the former, i.e., $\arg\max_j p^{-*}_{i,j}(u) = \hat{y}_u$. In this situation, just by using the cross-entropy term in Eq. 10, FedAPS could potentially forfeit the valuable information within the unselected model. Inspired by FedLabel [19], we employ a local-global consistency regularization term to handle scenarios where the unselected model predicts the identical label as the selected model, thus exploiting the information from the unselected model, as follows:

$$\begin{gathered} \Phi_{KL}(u) = \lambda_k(u) \cdot KL(p(w_{L,k}, u), p^{-*}(u)) \\ \text{s.t. } \arg\max_{j} p^{-*}_{i,j}(u) = \hat{y}_u, \end{gathered} \tag{11}$$

$$\lambda_k(u) := \lambda_0 \cdot s(p^{-*}(u))/s(p^*(u)), \tag{12}$$

where $\lambda_k(u)$, the ratio of the confidence scores of the non-selected model to the selected model, weights the local-global consistency regularization term (Kullback-Leibler divergence), i.e., $\lambda_0 \cdot s(p^{-*}(u))/s(p^*(u)) \leq \lambda_0$.

Therefore, the regularization weight is based on the unselected model's confidence score relative to the chosen model; the lower ratio means the smaller weight. $\lambda_k(u), u \in \mathcal{D}_{U,k}$ reaches its maximum value, λ_0, when the confidence scores of both the selected and non-selected models are equal.

Final Semi-Supervised Loss: By integrating the cross-entropy loss of confidence-based selection from Eq. 10 with the local global consistency regularity loss in Eq. 11, we can determine the final semi-supervised loss of FedAPS for each client $k \in \mathcal{C}^t$ on unlabeled data $\mathcal{D}_{U,k}$, denoted by $\mathcal{L}_{U,k}$, as following:

$$\Phi_{U,k} = \frac{1}{|\mathcal{D}_{U,k}|} \sum_{u \in \mathcal{D}_{U,k}} (\Phi_{CE}(u) + \Phi_{KL}(u)). \tag{13}$$

Now that we possess the ultimate semi-supervised loss of FedAPS in Eq. 13, we may further illustrate the implementation of FedAPS in the following subsections.

C. FedAPS Implementation

Perform training on unlabeled data:

$$w^{(t,\tau')}_{\mathcal{L},k} = w^{(t,0)}_{\mathcal{L},k} - \eta \nabla \Phi_{U,k}. \tag{14}$$

Semi-supervised for local updates:

$$\Delta w^t_U = w^{(t,\tau')}_{\mathcal{L},k} - w^{(t,0)}_{\mathcal{L},k}. \tag{15}$$

The cross-entropy loss on labeled data for FedAPS can be expressed as:

$$\Phi_{\mathcal{L},k} = \frac{1}{|\mathcal{D}_{\mathcal{L},k}|} \sum_{\xi \in \mathcal{D}_{\mathcal{L},k}} CrossEntropy(w^{(t,\tau')}_{\mathcal{L},k}, \xi), \tag{16}$$

where $\xi := \{x, y\}, \xi \in \mathcal{D}_{\mathcal{L},k}$.

Perform training on labeled data:

$$w^{(t,\tau)}_{\mathcal{L},k} = w^{(t,\tau')}_{\mathcal{L},k} - \eta \nabla \Phi_{\mathcal{L},k}. \tag{17}$$

Supervised for local updates:

$$\Delta w^t_{\mathcal{L}} = w^{(t,\tau)}_{\mathcal{L},k} - w^{(t,\tau')}_{\mathcal{L},k}. \tag{18}$$

The details of FedAPS's implementation is in Algorithm 1 and Algorithm 2.

IV. Experiments and Evaluation Results

In this section, we comprehensively evaluate the FedAPS framework on publicly available real-world datasets for air writing recognition tasks. The experiments aim to demonstrate: 1) the effectiveness of multi-modal data augmentation of FedAPS in varying label availability proportions; 2) the effectiveness of various modules of FedAPS; and 3) the robustness of FedAPS under different settings.

A. Experimental Setup

We will discuss the dataset, pre-processing, comparison methods, etc., in this part.

This article has been accepted for publication in IEEE Transactions on Consumer Electronics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TCE.2024.3467125

Algorithm 1 FedAPS Algorithm: Server Side

Input: communication rounds T, client number N;
Output: global model $w_{\mathcal{G}}^{T}$

procedure SERVER AGGREGATOR
  Initialize model $w_{\mathcal{G}}^{0}$ randomly
  for each round $t = 1, \ldots, T$ **do**
    Randomly select set $\mathcal{C}^t$ from N clients
    Distribute $w_{\mathcal{G}}^{t}$ to clients in $\mathcal{C}^t$
    for each client $k \in \mathcal{C}^t$ **in parallel do**
      $w_{\mathcal{L},k}^{(t,\tau)} \leftarrow$ **ClientUpdate** $(w_{\mathcal{L},k}^{(t,0)}, \mathcal{D}_k)$
      Upload local model $w_{\mathcal{L},k}^{(t,\tau)}$ to server
    end for
    $w_{\mathcal{G}}^{t+1} \leftarrow \frac{1}{|\mathcal{C}^t|} \sum_{k \in \mathcal{C}^t} w_{\mathcal{L},k}^{(t,\tau)}$ (via Aggregator)
  end for
end procedure

Algorithm 2 FedAPS Algorithm: Client Side

Input: global model $w^{(t,0)}$, local iterations τ
Output: local model $w_{\mathcal{L},k}^{(t,\tau)}$

procedure CLIENT LOCAL UPDATE
  Initialize model $w_{\mathcal{L},k}^{(t,0)} \leftarrow w^{(t,0)}$
  Set $w_{\mathcal{G},k}^{(t,0)} \leftarrow w^{(t,0)}$, $w_{\mathcal{H},k}^{(t,0)} \leftarrow w_{\mathcal{L},k}^{(t-1,\tau)}$
  // Adaptive pseudo-label selection
  Get $\hat{y}_u, u \in \mathcal{D}_{U,k}$ via Eq. 9
  // Semi-supervised training
  Get $w_{\mathcal{L},k}^{(t,\tau')} \leftarrow w_{\mathcal{L},k}^{(t,0)} + \Delta w_{U}^{t}$ (see 14 and 15)
  // Supervised training
  Get $w_{\mathcal{L},k}^{(t,\tau)} \leftarrow w_{\mathcal{L},k}^{(t,\tau')} + \Delta w_{\mathcal{L}}^{t}$ (see 17 and 18)
  Save local model $w_{\mathcal{L},k}^{(t,\tau)}$ of current round
  Upload local model $w_{\mathcal{L},k}^{(t,\tau)}$ to server
end procedure

1) Dataset and Processing: In order to assess the effectiveness of our proposed FedAPS framework, we chose to utilize the IAHEW-UCAS2016 [11] dataset, as it contains extensive recordings of actual air writing English words. Comprising recordings from 324 participants, the dataset encompasses a total of 150,480 instances, spanning 2,280 frequently used English words, each with 66 distinct samples.

Given that the IAHEW-UCAS2016 dataset is a sequence of fingertip trajectories collected by the Leap Motion sensors, we pre-processed each word recording to produce the binary image of size 100×32, where the trajectory point's position is one, and the other pixels are zero, as depicted in Fig. 3. Additionally, for the sake of simplicity, exclusively lowercase words were taken into account, while four specific words ("old-fashioned", "I", "Christmas", and "English") were ruled out [1]. Within each of the 66 distinct samples of every word, 60 samples were randomly allocated for training purposes, encompassing 136,560 recordings, while the remaining were designated for testing, comprising 13,656 recordings.

2) Comparison Approaches: We compare FedAPS with the following widely used methods:

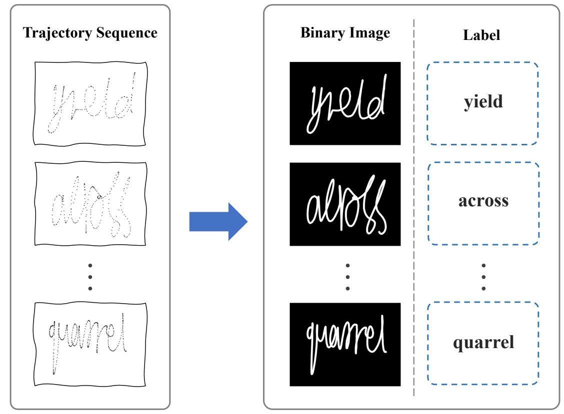


Fig. 3. An illustration of pre-processed IAHEW-UCAS2016 data converted from a sequence of fingertip coordinates to a binary image.

- **FixMatch** [20] is an advanced version of Pseudo-label [34]. It filters noisy unlabeled predictions using a fixed threshold and treats high-confidence predicted samples as pseudo-labels.
- **Dash** [22] improved FixMatch by proposing using dynamic thresholds rather than fixed thresholds to select more efficient unlabeled samples for training.
- **FreeMatch** [35] adaptively adjusts the confidence threshold depending on the learning status of the model and incorporates a self-adaptive class fairness regularization penalty to achieve diverse predictions in the early stages of training.
- **FedAvg** [36] is a pioneering work in federated learning. It merges information across devices by averaging local model parameters to learn a shared model globally without exposing their respective local data.
- **FedProx** [37] is a variation of the FedAvg algorithm. By introducing a proximal term, it controls the difference between client model updates and global model parameters. It can effectively alleviate the Non-IID challenge in traditional FL.
- **FedMatch** [32] combines inter-client consistency loss and parameter decomposition for semi-supervised learning in federated scenarios.
- **FedTriNet** [33] employs three networks and dynamic quality control to generate high-quality pseudo-labels for unlabeled data, incorporating them into the training set.
- **FedLabel** [19] to choose between the local and global models to generate pseudo-labels for unlabeled data.

The above methods can be conveniently grouped into three categories of baselines:

a) The SFL baselines with fully labeled data, which are the hypothetical upper bound that all data local of the client is labeled: **FedAvg (100%)** and **FedProx (100%)**
b) The SFL baselines with partially labeled data, which are the lower bound on the practical setting of FL, where only a few labeled data are available: **FedAvg** and **FedProx**.
c) The SSFL baselines with partially labeled data, which are advanced approaches to overcome the challenge of label

TABLE I
PERFORMANCE COMPARISON OF VARIOUS METHODS AT DIFFERENT LABEL RATIOS.

Approaches		Proportion of Labeled Data 5%	20%	50%
Supervised, Fully Labeled	FedAvg (100%)		93.37 (100%)	
	FedProx (100%)		93.92 (100%)	
Supervised, Partially Labeled	FedAvg	57.35	80.37	88.11
	FedProx	58.87	81.78	89.34
Semi-supervised, Partially Labeled	FedAvg + FixMatch	65.34	83.80	88.35
	FedAvg + Dash	69.08	85.53	89.74
	FedAvg + FreeMatch	70.34	85.91	89.84
	FedProx + FixMatch	67.69	84.26	89.11
	FedProx + Dash	71.25	86.01	89.05
	FedProx + FreeMatch	72.16	86.85	90.32
	FedMatch	72.99	85.53	90.16
	FedTriNet	66.30	82.68	89.23
	FedLabel	72.41	85.45	90.50
	FedAPS(ours)	**79.27**	**90.77**	**92.91**

scarcity in FL: **FedAvg+FixMatch**, **FedProx+FixMatch**, **FedAvg+Dash**, **FedProx+Dash**, **FedAvg+FreeMatch**, **FedProx+FreeMatch**, **FedMatch**, **FedTriNet** and **FedLabel**.

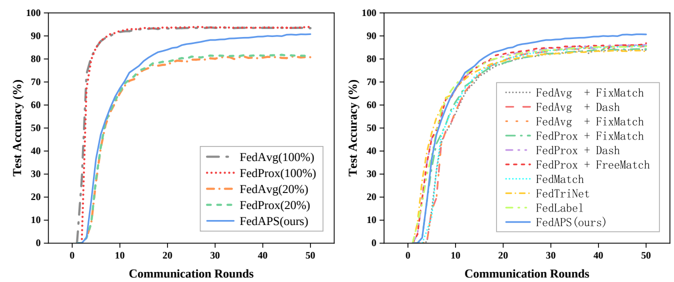
Note that our comparative analysis excludes approaches that introduce specific assumptions, such as certain clients having entirely labeled datasets or the server possessing labeled data.

3) Implementation Details: As the local backbone model, we adapt the TMBA model proposed by FedAWR [8], which is explicitly tailored for air writing recognition tasks. The AdaDelata optimizer [38] is used with a decay rate of 0.95, combined with a gradient clipping of magnitude 5.

In the federated setting, we considered 3 client devices as the default configuration and executed 50 rounds of federated communication, culminating in achieving a converged state. Each client device engaged in 4 epochs of local training with the batch size 16. We then set the selection threshold δ_1 to 0.95, with the recommendation threshold δ_2 to 0.85 and the regularization weight λ_0 to 1.

To ensure the reliability of the experiments, we repeated each experiment five times with distinct random initialization seeds. The final presentation of the results reflects the average performance derived from these five sets of experiments.

The training of all models transpired on the Nvidia GeForce RTX 3090 24GB GPU within the Ubuntu 20.04.1x64 operating system, complemented by a 16-core Intel CPU and 187GB RAM. The experimental setup harnessed Python version 3.8.13 and PyTorch version 1.13.0. In addition, the Flower [39] framework was employed to emulate the federated environment.



(a) Comparison with supervised baselines (b) Comparison with semi-supervised baselines

Fig. 4. Comparison of test accuracy for 20% labeled data proportion. FedAPS outperforms the advanced federated semi-supervised baselines and achieves performance comparable to fully labeled federated supervised learning.

B. Experimental Results

1) Performance of FedAPS under Different Label Ratios: In Table I and Fig. 4, we present the performance of FedAPS across various labeled data proportions. Specifically, we consider cases with 5% and 20% labeled data, referred to as low labeled proportion cases, and 50% labeled data as high labeled proportion cases. Compared to alternative semi-supervised federated learning approaches, FedAPS consistently achieves the highest test accuracy. Remarkably, its performance approaches that of supervised federated learning, even when the latter utilizes 100% labeled data.

It is noteworthy that, for scenarios with a high percentage of labeled data (50%), the performance gap between FedAPS and other baselines is relatively modest, typically between 3-5%. Conversely, in lower labeled data proportions (5% and 20%), FedAPS demonstrates a substantial performance superiority, outperforming other baselines by approximately 4-14%. This observed trend underscores FedAPS's pronounced advantage in scenarios characterized by a scarcity of labeled data. Thus, our results suggest that FedAPS excels, particularly in scenarios where labeled data is limited, showcasing its efficacy in overcoming challenges caused by high label scarcity.

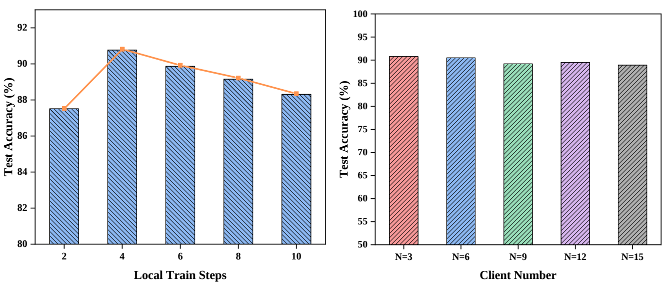
2) Robustness of FedAPS on Different Data Distributions: Table II shows the comparison of the performance of the various methods in IID and Non-IID scenarios. It is clear from the table that FedAPS outperforms the other baselines in both

TABLE II
PERFORMANCE COMPARISON OF DIFFERENT METHODS WITH DATA HETEROGENEITY.

Approaches	IID	Non-IID
FedAvg(100%)	93.95	93.70
FedProx(100%)	92.86	93.92
FedAvg	82.65	81.37
FedProx	81.13	81.78
FedAvg + FixMatch	84.60	83.80
FedAvg + Dash	85.59	85.53
FedAvg + FreeMatch	86.08	85.91
FedProx + FixMatch	85.23	84.26
FedProx + Dash	86.72	86.01
FedProx + FreeMatch	86.94	<u>86.85</u>
FedMatch	<u>87.22</u>	85.53
FedTriNet	84.65	82.68
FedLabel	84.97	85.45
FedAPS(ours)	**89.85**	**90.77**

TABLE III
PERFORMANCE COMPARISON OF DIFFERENT PSEUDO-LABELING METHODS.

Pseudo-labeling Method	Proportion of labeled data	
	5%	50%
Only Global	67.04	86.72
Only Historical Local	63.81	89.21
Historical Local + Global	72.69	90.85
FedAPS(ours)	79.27	92.91



(a) The impact of local train steps (b) The impact of client number

Fig. 5. The effectiveness of FedAPS diverse setting.

the IID and Non-IID identically distributed cases. This reflects the robustness of FedAPS to data heterogeneity. However, in the case of IID, the difference between FedAPS and the other baselines is insignificant, and the performance is closer. This means that in the presence of the data heterogeneity problem, the other baselines underperform that FedAPS can achieve.

3) Effectiveness of FedAPS across Diverse Settings: In order to further analyze FedAPS, we investigate its performance under various parameter settings in this subsection.

Varying Local Train Steps: After conducting semi-supervised training on unlabeled data, each client undergoes τ rounds of supervised training using locally labeled data to fine-tune the model. A larger τ theoretically implies a better reflection of the locally labeled data by the local model. Therefore, an experiment was conducted to investigate the impact of τ on the performance of FedAPS, as illustrated in Fig. 5(a).

The results depicted in Fig. 5(a) reveal that the optimal performance is achieved when τ is set to 4. Conversely, the poorest performance is observed with τ set to 2, indicating that an excessively small τ leads to insufficient model learning, preventing an adequate reflection of the locally labeled data. Additionally, performance improvement tends to saturate or slightly decline as τ increases beyond an optimal point. This phenomenon suggests that an excessively large τ may cause over-fitting of the local model due to the limited amount of locally labeled data, thereby restricting the model's generalization. Hence, achieving a balance is crucial. A moderately sized value for τ is essential to enable the local model to effectively reflect the client's local data and to extract maximum benefit from the global model without succumbing to overfitting limitations.

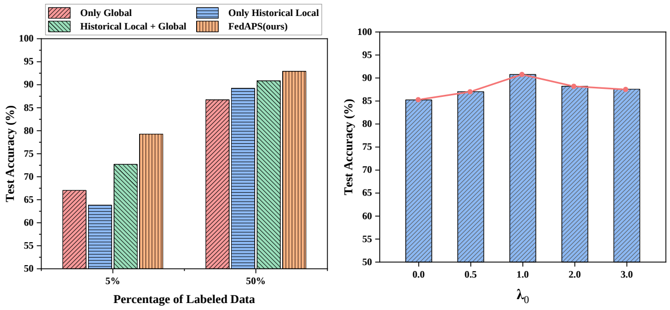
Varying Number of Clients: The number of clients N plays a crucial role in the FL process, with influences related to the data distribution and, thus, the generalizability of the global model. Specifically, introducing additional clients may exacerbate the problem of Non-IID among clients due to random data division. By conducting this ablation experiment, we aim to provide insight into the effectiveness of FedAPS in gradually increasing the number of clients. With a labeled data proportion of 20%, we gradually increased the number of clients N from 3 to 15 while keeping other parameters constant. The experimental results are shown in Fig. 5(b), where we observe that the impact of the number of clients on the performance of FedAPS is relatively small, and the overall performance remains at a comparable level without a significant performance degradation trend. With a gradual increase in the number of clients, FedAPS shows relatively robust performance and is not negatively affected by the introduction of too many clients. This result demonstrates the robustness of FedAPS to a different numbers of clients and its strong generalization ability in the face of non-IID scenarios.

4) Ablation Study: In this subsection, we focus on the effectiveness of two critical designs of FedAPS, i.e., adaptive pseudo-label selection and historical local-global consistency regularization terms.

Effectiveness of Adaptive Pseudo-label Selection: As can be observed from Table III and Fig. 6(a), "only global" performs better with the smaller proportion of labels, while "only historical local" exhibits better performance with the larger proportion of labels. The pseudo-labeling approach, which combines historical local and global models, outperforms using a single model for pseudo-labeling across different label ratios. However, in cases where both models fail to provide optimal pseudo-labels, there is a consequential loss of valuable unlabeled data.

Comparatively, the three pseudo-labeling methods' overall performance is inferior compared to our proposed FedAPS adaptive pseudo-labeling selection method for different label

This article has been accepted for publication in IEEE Transactions on Consumer Electronics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TCE.2024.3467125



(a) The effectiveness of adaptive pseudo-label selection

(b) The impact of λ_0

Fig. 6. Ablation study on the effect of λ_0 and the effect of adaptive pseudo-label selection.

ratios. This is attributed to the effectiveness of the dynamic word score recommendation strategy of FedAPS, which integrates dynamic and static features of the air writing data. This integration allows for more rational pseudo-labels when the prediction confidence of both models falls below a predetermined threshold. These results not only highlight the robustness of our proposed adaptive pseudo-label selection method under varying label ratios, but also emphasise the effectiveness of the dynamic word score recommendation strategy in pseudo-labeling.

Effectiveness of Consistency Regularization: In our FedAPS framework, we incorporate a historical local-global consistency regularization term designed to capture valuable information from unselected models, with its weight determined by the parameter λ_0. The results are shown in Fig. 6(b), where ablation experiments on λ_0 were performed to assess the effectiveness of this regularization term. When λ_0 is set to 0, indicating the absence of the regularization term, the test accuracy reaches its lowest point. This suggests that the decision between the historical local model and the global model may result in the loss of valuable knowledge contained in the unselected model.

As λ_0 is increased, a noticeable enhancement in test accuracy is observed. However, we observe a diminishing return on improvement beyond a certain threshold. Consequently, adjusting the degree of historical local-global consistency regularization can be achieved by fine-tuning λ_0 to an optimal value, ensuring the maximum regularization gain. This operation allows for preserving valuable knowledge from unselected models while avoiding excessive regularization that may hinder overall performance.

V. Conclusion

In conclusion, the FedAPS framework marks an advancement in federated semi-supervised learning, particularly for air writing recognition in consumer electronics. It uniquely enables clients to make data-driven decisions by adaptively selecting the optimal model for multi-modal data augmentation through pseudo-labeling from historical local models, global models, or dynamic word score recommendations. This approach exemplifies multi-modal data-driven decision-making, crucial for next-generation technologies. The inclusion of a historical local-global consistency regularization term further enhances decision accuracy by utilizing knowledge from historical models. Our evaluations on a real-world air writing dataset demonstrate FedAPS's superiority over existing baselines and its comparable performance to fully supervised methods despite limited multi-modal data. These findings affirm its effectiveness in addressing practical challenges in air writing recognition while enhancing privacy preservation.

For future research, we plan to explore automated methods for improving adaptive pseudo-labeling selection for higher performance of multi-modal data augmentation. Besides, more complex multi-modal data input scenarios should be taken into consideration.

References

[1] S.-K. Lee and J.-H. Kim, "Air-text: Air-writing and recognition system," in *Proceedings of the 29th ACM International Conference on Multimedia*. Association for Computing Machinery, 2021, pp. 1267–1274.

[2] T. He, J. Chen, B.-G. He, W. Wang, Z.-l. Zhu, and Z. Lv, "Towards the wearable sensors: Advances, trends and challenges," *ACM Computing Surveys*, vol. 55, no. 14s, 2023.

[3] M. Fukui, T. Watanabe, and M. Kanazawa, "Sound source separation for plural passenger speech recognition in smart mobility system," *IEEE Transactions on Consumer Electronics*, vol. 64, no. 3, pp. 399–405, 2018.

[4] S. M. Massa, D. Riboni, and K. Nazarpour, "Explainable ai-powered graph neural networks for hd emg-based gesture intention recognition," *IEEE Transactions on Consumer Electronics*, vol. 70, no. 1, pp. 4499–4506, 2024.

[5] X. Kong, J. Wang, Z. Hu, Y. He, X. Zhao, and G. Shen, "Mobile trajectory anomaly detection: Taxonomy, methodology, challenges, and directions," *IEEE Internet of Things Journal*, vol. 11, no. 11, pp. 19 210–19 231, 2024.

[6] U.-H. Kim, S.-M. Yoo, and J.-H. Kim, "I-keyboard: Fully imaginary keyboard on touch devices empowered by deep neural decoder," *IEEE Transactions on Cybernetics*, vol. 51, no. 9, pp. 4528–4539, 2021.

[7] J.-H. Kim, S.-J. Choi, and J.-W. Jeong, "Watch & do: A smart iot interaction system with object detection and gaze estimation," *IEEE Transactions on Consumer Electronics*, vol. 65, no. 2, pp. 195–204, 2019.

[8] X. Kong, W. Zhang, Y. Qu, X. Yao, and G. Shen, "Fedawr: An interactive federated active learning framework for air writing recognition," *IEEE Transactions on Mobile Computing*, vol. 23, no. 5, pp. 6423–6436, 2024.

[9] U.-H. Kim, Y. Hwang, S.-K. Lee, and J.-H. Kim, "Writing in the air: Unconstrained text recognition from finger movement using spatio-temporal convolution," *IEEE Transactions on Artificial Intelligence*, vol. 4, no. 6, pp. 1386–1398, 2023.

[10] Y. Gao, S. Zeng, J. Zhao, W. Liu, and W. Dong, "Airtext: One-handed text entry in the air for cots smartwatches," *IEEE Transactions on Mobile Computing*, vol. 22, no. 5, pp. 2506–2519, 2023.

[11] J. Gan and W. Wang, "In-air handwritten english word recognition using attention recurrent translator," *Neural Computing and Applications*, vol. 31, pp. 3155–3172, 2019.

[12] J. Gan, W. Wang, and K. Lu, "A unified cnn-rnn approach for in-air handwritten english word recognition," in *2018 IEEE International Conference on Multimedia and Expo (ICME)*. IEEE, 2018, pp. 1–6.

[13] ——, "In-air handwritten chinese text recognition with temporal convolutional recurrent network," *Pattern Recognition*, vol. 97, p. 107025, 2020.

[14] X. Kong, H. Yuan, G. Shen, H. Zhou, W. Liu, and Y. Yang, "Mitigating data imbalance and generating better prototypes in heterogeneous federated graph learning," *Knowledge-Based Systems*, vol. 296, p. 111876, 2024.

[15] L. Wang, H. Zhou, Y. Bao, X. Yan, G. Shen, and X. Kong, "Horizontal federated recommender system: A survey," *ACM Computing Surveys*, vol. 56, no. 9, 2024.

[16] P. Liu, L. Jiang, H. Lin, J. Hu, S. Garg, and M. Alrashoud, "Federated multimodal learning for privacy-preserving driver break recommendations in consumer electronics," *IEEE Transactions on Consumer Electronics*, vol. 70, no. 1, pp. 4564–4573, 2024.

[17] X. Kong, W. Zhang, H. Wang, M. Hou, X. Chen, X. Yan, and S. K. Das, "Federated graph anomaly detection via contrastive self-supervised learning," *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–14, 2024.

[18] Y. Jin, Y. Liu, K. Chen, and Q. Yang, "Federated learning without full labels: A survey," *arXiv preprint arXiv:2303.14453*, 2023.

[19] Y. J. Cho, G. Joshi, and D. Dimitriadis, "Local or global: Selective knowledge assimilation for federated learning with limited labels," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2023, pp. 17 087–17 096.

[20] K. Sohn, D. Berthelot, N. Carlini, Z. Zhang, H. Zhang, C. A. Raffel, E. D. Cubuk, A. Kurakin, and C.-L. Li, "Fixmatch: Simplifying semi-supervised learning with consistency and confidence," in *Advances in Neural Information Processing Systems*, H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan, and H. Lin, Eds., vol. 33. Curran Associates, Inc., 2020, pp. 596–608.

[21] B. Zhang, Y. Wang, W. Hou, H. Wu, J. Wang, M. Okumura, and T. Shinozaki, "Flexmatch: Boosting semi-supervised learning with curriculum pseudo labeling," in *Advances in Neural Information Processing Systems*, vol. 34, 2021, pp. 18 408–18 419.

[22] Y. Xu, L. Shang, J. Ye, Q. Qian, Y.-F. Li, B. Sun, H. Li, and R. Jin, "Dash: Semi-supervised learning with dynamic thresholding," in *Proceedings of the 38th International Conference on Machine Learning*, vol. 139. PMLR, 2021, pp. 11 525–11 536.

[23] Q. Xie, Z. Dai, E. Hovy, T. Luong, and Q. Le, "Unsupervised data augmentation for consistency training," in *Advances in Neural Information Processing Systems*, vol. 33, 2020, pp. 6256–6268.

[24] J. Jeong, S. Lee, J. Kim, and N. Kwak, "Consistency-based semi-supervised learning for object detection," in *Advances in Neural Information Processing Systems*, vol. 32, 2019.

[25] G. Li, X. Li, Y. Wang, Y. Wu, D. Liang, and S. Zhang, "Pseco: Pseudo labeling and consistency training for semi-supervised object detection," in *European Conference on Computer Vision*. Springer, 2022, pp. 457–472.

[26] J.-M. Guo, C.-C. Sun, K.-Y. Chan, and C.-Y. Liu, "Robust semi-supervised learning with multi-consistency and data augmentation," *IEEE Transactions on Consumer Electronics*, vol. 70, no. 1, pp. 414–424, 2024.

[27] C. Zhang, F. Wu, J. Yi, D. Xu, Y. Yu, J. Wang, Y. Wang, T. Xu, X. Xie, and E. Chen, "Non-iid always bad? semi-supervised heterogeneous federated learning with local knowledge enhancement," in *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*. Association for Computing Machinery, 2023, pp. 3257–3267.

[28] E. Diao, J. Ding, and V. Tarokh, "Semifl: Communication efficient semi-supervised federated learning with unlabeled clients," *arXiv preprint arXiv:2106.01432*, 2021.

[29] H. Yu, Z. Chen, X. Zhang, X. Chen, F. Zhuang, H. Xiong, and X. Cheng, "Fedhar: Semi-supervised online learning for personalized federated human activity recognition," *IEEE Transactions on Mobile Computing*, vol. 22, no. 6, pp. 3318–3332, 2023.

[30] Z. Long, J. Wang, Y. Wang, H. Xiao, and F. Ma, "Fedcon: A contrastive framework for federated semi-supervised learning," *arXiv preprint arXiv:2109.04533*, 2021.

[31] M. Li, Q. Li, and Y. Wang, "Class balanced adaptive pseudo labeling for federated semi-supervised learning," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023, pp. 16 292–16 301.

[32] J. Chen, R. Zhang, J. Guo, Y. Fan, and X. Cheng, "Fedmatch: Federated learning over heterogeneous question answering data," in *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*. Association for Computing Machinery, 2021, pp. 181–190.

[33] L. Che, Z. Long, J. Wang, Y. Wang, H. Xiao, and F. Ma, "Fedtrinet: A pseudo labeling method with three players for federated semi-supervised learning," in *2021 IEEE International Conference on Big Data (Big Data)*. IEEE, 2021, pp. 715–724.

[34] D.-H. Lee *et al.*, "Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks," in *Workshop on challenges in representation learning (ICML)*, vol. 3, no. 2. Atlanta, 2013, p. 896.

[35] Y. Wang, H. Chen, Q. Heng, W. Hou, Y. Fan, Z. Wu, J. Wang, M. Savvides, T. Shinozaki, B. Raj *et al.*, "Freematch: Self-adaptive thresholding for semi-supervised learning," in *The Eleventh International Conference on Learning Representations*, 2023. [Online]. Available: https://openreview.net/forum?id=PDrUPTXJI_A

[36] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y. Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, vol. 54. PMLR, 2017, pp. 1273–1282.

[37] T. Li, A. K. Sahu, M. Zaheer, M. Sanjabi, A. Talwalkar, and V. Smith, "Federated optimization in heterogeneous networks," *Proceedings of Machine Learning and Systems*, vol. 2, pp. 429–450, 2020.

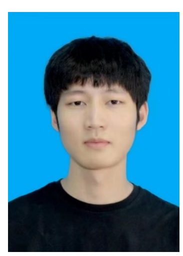
[38] J. Baek, G. Kim, J. Lee, S. Park, D. Han, S. Yun, S. J. Oh, and H. Lee, "What is wrong with scene text recognition model comparisons? dataset and model analysis," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2019, pp. 4715–4723.

[39] D. J. Beutel, T. Topal, A. Mathur, X. Qiu, T. Parcollet, P. P. de Gusmão, and N. D. Lane, "Flower: A friendly federated learning research framework," *arXiv preprint arXiv:2007.14390*, 2022.

Wenyi Zhang received the B.Sc. degree from the College of Science, Jiangxi University of Science and Technology, Jiangxi, China, in 2020. He is currently pursuing a Ph.D. degree in Computer Science and Technology with the School of Computer Science, Zhejiang University of Technology, China. His research interests include gesture recognition, mobile edge computing, and federated learning.

Xiangjie Kong (IEEE Senior Member) received the B.Sc. and Ph.D. degrees from Zhejiang University, Hangzhou, China, in 2004 and 2009, respectively. He is a Professor with College of Computer Science and Technology, Zhejiang University of Technology, China. Previously, he was an Associate Professor with the School of Software, Dalian University of Technology, China. He has published over 180 scientific papers in international journals and conferences (with over 150 indexed by ISI SCIE). His research interests include social computing, mobile computing, and computational social science. He is a Senior Member of the IEEE, a Distinguished Member of CCF, and is a member of ACM.

Guojiang Shen received the B.Sc. degree in Control Theory and Control Engineering and the Ph.D. degree in Control Science and Engineering from Zhejiang University, Hangzhou, China, in 1999 and 2004, respectively. He is currently a Professor in the College of Computer Science and Technology, Zhejiang University of Technology. His current research interests include artificial intelligence, Big Data analytics and intelligent transportation systems.



Zhenpeng Wu received the B.E. degree in Internet of Things from Zhejiang University of Technology in 2023. He is currently pursuing for the ME degree in Electronic and Information Science at the School of Computer Science, Zhejiang University of Technology, China. His research interests focus on gesture recognition and federal learning.

Youyang Qu (IEEE Member) is currently a research scientist of data61, Commonwealth Scientific and Industrial Research Organization (CSIRO), Australia. Before joining CSIRO, he served as a research fellow at Deakin University. He received his B.S. degree of Mechanical Automation in 2012 and M.S. degree in Software Engineering in 2015 from Beijing Institute of Technology, respectively. He received his PhD degree at the School of Information Technology, Deakin University in 2019. His research interests focus on Machine Learning, Big Data, IoT, blockchain and corresponding security and customizable privacy issues. He has over 50 publications, including high-quality journals and conferences papers such as IEEE TII, IEEE TNSE, ACM Computing Surveys, IEEE IOTJ, etc. He is active in the research society and has served as an organizing committee member in SPDE 2020, BigSecuirty 2021, Tridentcom 2021/2022.